93
B | IFW

Docket No.: 4441-0106P
(PATENT)

OIPE IAP45
FEB 08 2006
PATENT & TRADEMARK OFFICE

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Fredrik HENN et al.

Application No.: 09/987,657

Confirmation No.: 7524

Filed: November 15, 2001

Art Unit: 2654

For: ENHANCING THE PERFORMANCE OF CODING SYSTEMS THAT USE HIGH FREQUENCY RECONSTRUCTION METHODS

Examiner: V. B. Chawan

CLAIM FOR PRIORITY AND SUBMISSION OF DOCUMENTS

HANDCARRY TO:
U.S. Patent and Trademark Office
Customer Service Window - MS ISSUE FEE
Randolph Building
401 Dulany Street
Alexandria, VA 22314

Sir:

Applicants hereby claim priority under 35 U.S.C. 119 based on the following prior foreign application filed in the following foreign country on the date indicated:

Country	Application No.	Date
Sweden	0004187-1	November 15, 2000

In support of this claim, a certified copy of the said original foreign application is filed herewith.

Dated: February 8, 2006

Respectfully submitted,

By Terrell C. Birch

Terrell C. Birch, #,19,382
BIRCH, STEWART, KOLASCH & BIRCH, LLP
8110 Gatehouse Road, Suite 100 East
P.O. Box 747
Falls Church, Virginia 22040-0747
(703) 205-8000
Attorney for Applicant

Attachment

F. HENN et al
09/987,657
filed 11/15/2001
Cust. # 02292
1 of 1
4441-0106D

PRV

PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen

Intyg
Certificate

PATENT- OCH REGISTRERINGSVERKET

Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.

(71) Sökande / Applicant (s) — *Coding Technologies Sweden AB, Stockholm SE*

(21) Patentansökningsnummer / Patent application number — *0004187-1*

(86) Ingivningsdatum / Date of filing — *2000-11-15*

Stockholm, 2006-01-24

För Patent- och registreringsverket
For the Patent- and Registration Office

Hjördis Segerlund

Hjördis Segerlund

Avgift
Fee 170:-

CERTIFIED COPY OF PRIORITY DOCUMENT

PATENT- OCH REGISTRERINGSVERKET
SWEDEN

Postadress/Adress
Box 5055
S-102 42 STOCKHOLM

Telefon/Phone
+46 8 782 25 00
Vx 08-782 25 00

Telex
17978
PATOREG S

Telefax
+46 8 666 02 86
08-666 02 86

ENHANCING THE PERFORMANCE OF CODING SYSTEMS THAT USE HIGH FREQUENCY RECONSTRUCTION METHODS

TECHNICAL FIELD

The present invention relates to a new method and apparatus for an adaptive crossover frequency between the range covered by a high frequency reconstruction method and an underlying base coder technology over time. The method may be used both for natural audio coding and speech coding and is especially suited for coders using SBR [WO 98/57436] or other high frequency reconstruction methods.

BACKGROUND OF THE INVENTION

Audio source coding techniques can be divided into two classes: natural audio coding and speech coding. Natural audio coding is commonly used for music or arbitrary signals at medium bit rates, and generally offers wide audio bandwidth. Speech coders are basically limited to speech reproduction but can on the other hand be used at very low bit rates, albeit with low audio bandwidth. In both classes, the signal is generally separated into two major signal components, the "spectral envelope" and the corresponding "residual" signal. Throughout the following description, the term "spectral envelope" refers to the coarse spectral distribution of the signal in a general sense, e.g. filter coefficients in a linear prediction based coder or a set of time-frequency averages of sub-band samples in a sub-band coder. The term "residual" refers to the fine spectral distribution in a general sense, e.g. the LPC error signal or sub band samples normalized using the above time-frequency averages. "Envelope data" refers to the quantized and coded spectral envelope, and "residual data" to the quantized and coded residual.

Prior art codecs that make use of such a division between spectral envelope and residual exploit the fact that the spectral envelope can be coded much more efficiently than the residual. Especially in the case where high frequency reconstruction methods are used, only the envelope data of the upper frequency range is transmitted. However, a fixed crossover frequency between this upper and the lower frequency range was used. The crossover frequency itself needed to be determined based on a well-balanced trade off between more efficient envelope coding versus the price of more perceptual distortions due to simplifying the signal representation.

SUMMARY OF THE INVENTION

It is highly desirable to balance the tradeoff on the best choice of the crossover frequency on runtime and adjust it over time rather than using a fixed cross over frequency. The present invention provides a new method, and an apparatus for improving coding systems where high frequency reconstruction methods (HFR) are used. This enhancement to the existing coding schemes is designed to meet the special requirements of systems, where the residual signal within certain frequency regions is excluded from the transmitted data. Examples are systems employing HFR (High Frequency Reconstruction), in particular SBR (Spectral Band Replication), or parametric coders. The new invention breaks the traditional concept of a fixed crossover frequency between frequency ranges where standard coding schemes and a HFR coding scheme is used by detecting the optimum choice of the crossover frequency based on several parameters and then applying exactly the crossover frequency which is optimal at a

given time. As input parameters for the crossover frequency detection algorithm for a example a workload measure based on a psychoacoustic model, a spectral tonality analysis, short time bit demand detection or combinations of these three can be used. Thus, applying a flexible crossover frequency results in a substantial improvement since the optimum choice changes frequently over time resulting in a more constant and improved audio quality that is less dependent on program material characteristics.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of illustrative examples, not limiting the scope or spirit of the invention, with reference to the accompanying drawings, in which:

Fig. 1 is a graph illustrating the term frequency range and crossover frequency.

Fig. 2 is a block diagram of an encoder using a HFR method on which the present invention is based on, enhanced by a crossover frequency detection module.

Fig. 3 is a block diagram of a corresponding decoder using a HFR method.

Fig. 4 is a block diagram, which illustrates the crossover frequency detection module in detail.

Fig. 5 is a graph that illustrates the practical use of a workload measure.

Fig. 6 is a graph that illustrates the short time bit-demand variations of a constant bit rate coder.

Fig. 7 is a spectral frequency graph to illustrate the tonality analysis.

DESCRIPTION OF PREFERRED EMBODIMENTS

The below-described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

In a system where the low band frequency range 101 as given in Fig. 1 is encoded by a core coder and the high band frequency range 102 is covered by a suitable HFR method we can define the border between the two ranges as the crossover frequency 103. In prior art systems the potential in straitening the two coding schemes together by applying a dynamic crossover frequency over time was not exploited. Since the encoding schemes operate on a block wise frame by frame basis one is free to adapt the crossover frequency for every processed frame. It is possible to set up an appropriate detection algorithm that is able to take a choice on the crossover frequency such that the optimum quality for the joint coding system is achieved.

Taking into account that the audio quality of the core coder is also the basis for the quality of the reconstructed high band, it is obvious that a good and constant audio quality in the low band range is desired. By lowering the crossover frequency, the low band range, which the core coder has to cope with, is smaller and thus easier to encode. Thus measuring the degree of difficulty of encoding a frame and adjusting the crossover frequency accordingly, a more constant audio quality of the core encoder can be achieved. As an example on how to measure the degree of difficulty, the perceptual entropy approach as introduced by Johnston et al. may be used: Here a psychoacoustic model based on a spectral analysis is applied. Usually the spectral lines of the analysis filter bank are grouped to bands, whereas the number of lines within a band depends on its frequency and is chosen according to the well known bark scale, aiming at a constant perceptual frequency resolution for all bands. Using a psychoacoustic model that exploits effects such as spectral or temporal masking, one obtains thresholds of audibility. Taking the number of lines within one band, the calculated threshold of audibility and its spectral energy, one obtains the perceptual entropy within the band by applying the formula

$$pe = 0.5 \cdot \sum_{i=0}^{width-1} \log\left(\frac{ratio(i)}{\log(2)}\right) + n$$

where

$$ratio(i) = s(i)^2 \cdot \frac{width}{thr}$$

i = spectral line index within current band
$width$ = number of lines in current band
n = the number of lines in current band for which for which $ratio(i) > 1.0$ is true
$s(i)$ = spectral value of line i
thr = psychoacoustic threshold for current band.

and only terms with $ratio(i) > 1.0$ are used in the summation.

By summing up the perceptual entropies of all bands that have to be coded in the low band frequency range, one obtains a measure for the encoding difficulty for the current frame.

A similar approach is to calculate the distortion energy at the end of the encoding process of the core encoder by summing up the distortion energy of every band, i.e.

$$n = \sum_{b=0}^{bands-1} n_{dist}(b)$$

where

b = band index
$bands$ = number of bands

and the distortion of band b is given by

$$n_{dist}(b) = \begin{cases} n_{quant}(b) - thr(b) & for \quad n_{quant}(b)/thr(b) > 1.0 \\ 0 & otherwise \end{cases}$$

where

$n_{quant}(b)$ = quantization noise energy within the band
$thr(b)$ = psychoacoustic threshold for the band

In both cases a high perceptual entropy, respectively a high distortion energy indicates a difficult signal which is hard to code and is very likely to produce audible artifacts in the low band. In this case the crossover frequency detection module shall signal to use a lower crossover frequency in order to make it easier for the perceptual audio encoder to cope with the given signal. Concurrently a low perceptual energy, respectively a low distorted energy indicates an easy to code signal. Thus the crossover frequency shall be chosen higher in order to allow a wider frequency range for the low band, in order to avoid as much artifacts that are likely to be introduced in the high band range due to the limited capabilities of any existing HFR method. Both approaches also allow to use an Analysis-by-Synthesis approach by re-encoding the current frame if an adjustment of the crossover frequency was carried out. However, since in most state-of-the-art audio codecs overlapping transforms are used the performance of the overall system may be further improved by applying a smoothing of the output parameters over time in order to avoid too frequent switching of the crossover frequency to avoid blocking effects.

Besides the encoding difficulty of the current frame, another important parameter to base the best choice of the crossover frequency on is described as follows: A large number of audio signals such as speech or some musical instruments show the property that the spectral range can be divided into a pitched or tonal range and a noise-like range. Using tonality and/or noise analysis methods in the spectral domain, one may detect two ranges whereas each can be classified as tonal respectively noise-like. Thus the crossover frequency between these ranges is used as the

crossover frequency in the context of the present invention in order to better separate the tonal and noise like spectral range and feed them separately to the core encoder respectively the HFR method. Hence the overall audio quality of the combined codec system can be substantially improved in such cases.

Practical Implementations

An example of the encoder side as known by prior art, forming the basis on which the invention is built, is shown in Fig. 2. The analogue input signal is fed to an A/D-converter 201, forming a digital signal. The digital audio signal is fed to a core encoder 202, where source coding is performed. In addition, the digital signal is fed to a HFR encoder 203. The output of the HFR encoder represents the encoded envelope data covering the high band frequency range 102 starting at the crossover frequency 103 as illustrated in Fig. 1. The number of bits that is needed for the envelope data in the HFR encoder is passed to the core encoder in order to be subtracted from the total available bits for a given frame. The core encoder will then encode the remaining low band frequency range up to the crossover frequency. Both encoded signals are then passed to the multiplexer 205, forming a serial bit stream that is transmitted or stored.

The corresponding decoder side is shown in Fig. 3. The demultiplexer 301 separates the signals and feeds the appropriate part to the core decoder 302, which produces the low band digital audio signal. The envelope data is fed from the demultiplexer to the HFR envelope decoder 303, which decodes the data into a representation of the spectral envelope for the high band frequency range. The decoded envelope data is then fed to the gain control module 304. The low band signal from the audio decoder is routed to the transposition module 305, which generates a replicated high band signal from the low band. The high band signal is fed to the gain control module in order to adjust the high band envelope shape to that of the transmitted envelope. The output is thus an envelope adjusted high band audio signal. This signal is added to the output from the delay unit 306, which is fed with the low band audio signal whereas the delay compensates for the processing time of the high band signal. Finally, the obtained digital wideband signal is converted to an analogue audio signal in the digital to analogue converter 307.

According to the present invention, a crossover frequency detection module 204 respectively 401 in Fig. 4, is added to the encoder processing chain in order to find the optimum choice of the crossover frequency, which thus is variable over time. In addition, the choice has to be signaled and transmitted to the decoder which is described below. In order to achieve the optimum choice several input parameters are taken into account as illustrated in Fig. 4: An encoder workload measure analysis module 402 explores how difficult the current frame is to code for the core encoder using for example the perceptual entropy or the distorted energy approach as described above. Fig. 5 gives an example of the distortion energy 501, and the corresponding workload measure 502 of an perceptual audio coder. It can be observed that the value shows high deviations over time and is dependent on the input material characteristics.

As an additional input parameter to the detection module in a practical implementation, the short time bit-demand variation analysis 403 in the case of a constant bit rate audio codec may be used. State-of-the-art audio encoders such as MPEG Layer-3 or MPEG-2 AAC use a bit reservoir technique in order to compensate for short time peak bit-demand deviations from the average number of available bits per frame. Checking the fullness of such a bit

reservoir is a valuable indication of whether the audio encoder is currently able to cope well with an upcoming difficult to encode frame or not. A practical example of the number of used bits per frame 601, and the reservoir fullness over time 602 is given in Fig. 6. Thus, if the bit reservoir fullness is high, the core encoder will be able to handle a difficult frame and there is no need to choose a lower crossover frequency. Concurrently, the resulting audio quality may be substantially improved in the following frames by lowering the crossover frequency in order to allow a low bit demand cost for the core encoder, wherby a low bit reservoir fullness can be filled up due to the smaller frequency range that has to be encoded.

Fig. 7 shows the spectrum of the digital audio input signal 701, and a tonality indication curve 702, whereas the tonality can be calculated as given for example in AAC-Standard ISO/IEC 13818-7:1997(E),pp. 96-98, section B.2.1.4 "Steps in threshold calculation". Other well known tonality or noise detection algorithms such as spectral flatness measure are also suited for the given purpose: for the given example it is obvious that a low band and a high band range can be identified that do have totally different character with regard to their tonality, respectively noisiness. Thus for such signals the output of the tonality analysis module 404 shall signal to use the crossover frequency which divides the two ranges with the different properties.

All three input parameters to the joint detection module 405 can be combined and tuned according to the actual implementation of the used core encoder and the HFR encoder in order to obtain the maximum overall performance.

The output of the crossover frequency detection module 401, respectively 204, in form of the optimum choice of the crossover frequency is fed to both the core encoder and the HFR method encoder in order to signal each of the encoders the frequency range that shall be encoded. Again the number of bits needed by the HFR encoder is passed to the core encoder to calculate the number of available bits. The frequency range for each of the two coding schemes is also encoded, for example by an efficient table lookup scheme. If the frequency range between two subsequent frames does not change, this can be signaled by one single bit in order to keep the bit rate overhead as small as possible whereas the frequency ranges do not have to be transmitted explicitly in every frame. The encoded data of both encoders as well as the encoded frequency range data is then fed to the multiplexer, forming a serial bit stream that is transmitted or stored.

CLAIMS

1. A method for improving the performance of a source coding system comprising of a core coder for coding of a lower frequency band reaching up to a crossover frequency, and a HFR system for coding of a higher frequency band starting at said crossover frequency, where said HFR system in the synthesis at a decoder is guided by envelope data corresponding to said higher frequency range, **characterised by**

in an encoder, for a given frame calculate the value of said crossover frequency that yields the best tradeoff between core coder and HFR artifacts, and

for said frame, transmit or store said envelope data together with a control signal that describes said value.

2. A method according to claim **1**, **characterised in** that said calculation is based on a measure of the degree of difficulty of encoding a signal with said core coder, and a high difficulty lowers said value, and correspondingly, a low difficulty increases said value.

3. A method according to claim **2**, **characterised in** that said measure is based on the perceptual entropy of a signal.

4. A method according to claim **2**, **characterised in** that said measure is based on the distortion energy after coding with said core coder.

5. A method according to claim **2**, **characterised in** that said measure is based on the status of a bit-reservoir associated with said core coder.

6. A method according to claim **2**, **characterised in** that a combination of perceptual entropy, core coder distortion, and core coder bit-reservoir status is used in said calculation.

7. A method according to claim **1**, **characterised in** that said calculation is based on detection of a change in properties versus frequency of an input signal, and said crossover frequency is selected close to the frequency of said change.

8. A method according to claim **7**, **characterised in** that said detection discriminates between noise-like and tonal signals.

9. A method according to claim **1**, **characterised in** that said selection is based on a combination of a measure of difficulty of encoding a signal with said core coder, and detection of a change in properties versus frequency of said signal.

10. A source coding system comprising of means for coding of a lower frequency band reaching up to a crossover frequency, and means for HFR for coding of a higher frequency band starting at said crossover frequency, where said HFR means in the synthesis at a decoder is guided by envelope data corresponding to said higher frequency range, **characterised by**

in an encoder, means for calculation of the value of said crossover frequency that for a given frame yields the best tradeoff between artifacts from said means for coding of said lower frequency band and said HFR means,
means for generation of said envelope data for said frame using said value,
means for transmission or storage of said envelope data together with a control signal that describes said value,
means for decoding of said frame, using said control signal and said envelope data.

11. A system according to claim **10, characterised in** that said means for calculation yields a measure of difficulty of encoding a signal with said means for coding of said lower frequency band, and a high difficulty lowers said crossover frequency, and correspondingly, a low difficulty increases said crossover frequency.

12. A system according to claim **10, characterised in** that said means for calculation comprises means for noise/tonality detection, and said crossover frequency is selected close to the onset of a noise-like upper frequency range.

PRV00-11-15

ABSTRACT

The present invention relates to audio and /or speech coding and in particular to a new method and an apparatus for a dynamic adaptation of the crossover frequency for high frequency reconstructions methods over time. The invention teaches how to detect and signal the optimum crossover frequency between the high frequency reconstruction method and the underlying base coder technology. Using this technique, an improvement of the audio quality in the core coder and a more constant and improved audio quality of the overall system can be achieved. The method is applicable to both natural audio coding and speech coding systems and is especially suited for coders using SBR [WO 98/57436] or other high frequency reconstruction methods.

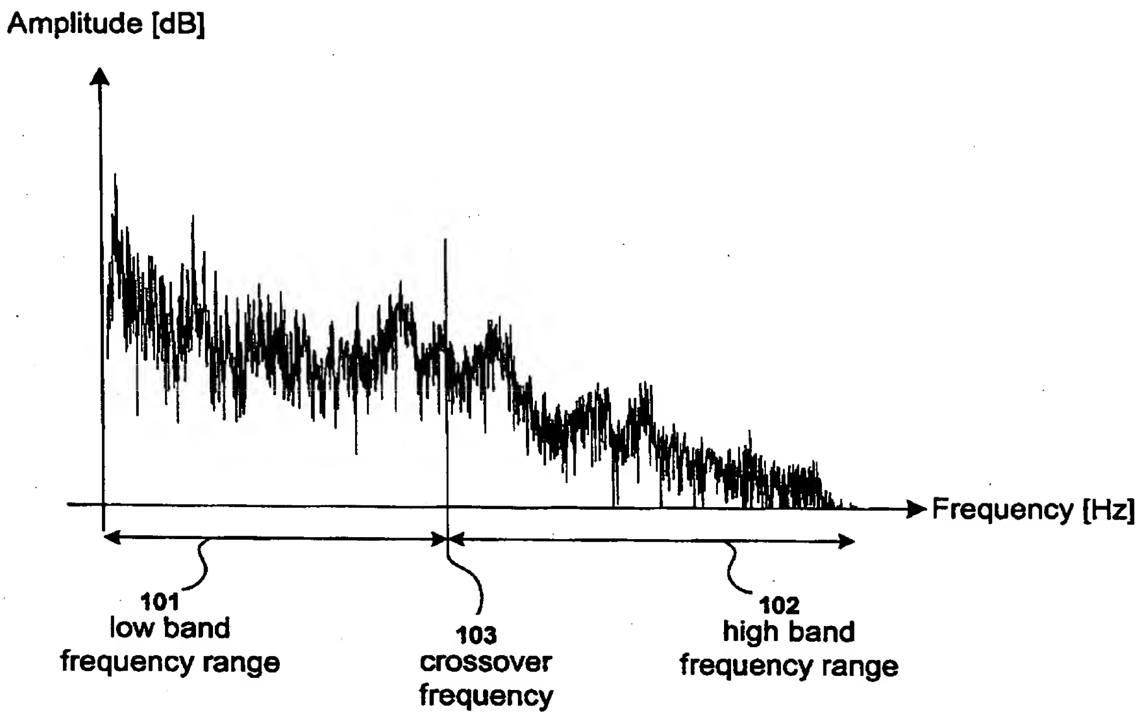

Amplitude [dB]
Frequency [Hz]
101
low band
frequency range
103
crossover
frequency
102
high band
frequency range

Fig. 1

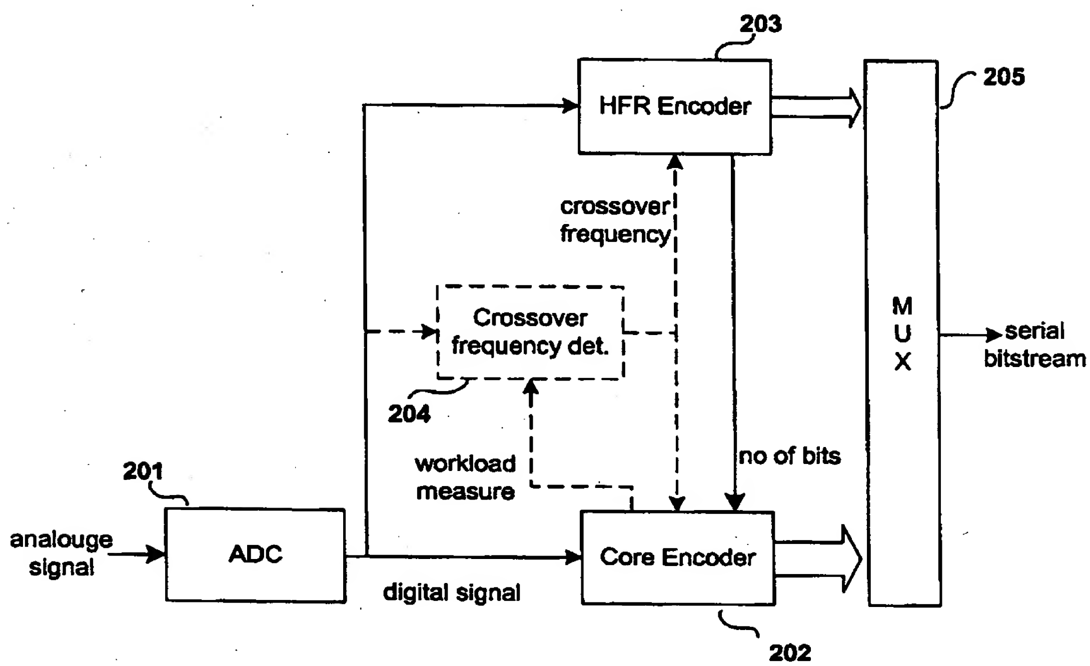

203
HFR Encoder
205
crossover
frequency
Crossover
frequency det.
204
M
U
X
serial
bitstream
201
workload
measure
no of bits
analouge
signal
ADC
Core Encoder
digital signal
202

Fig. 2

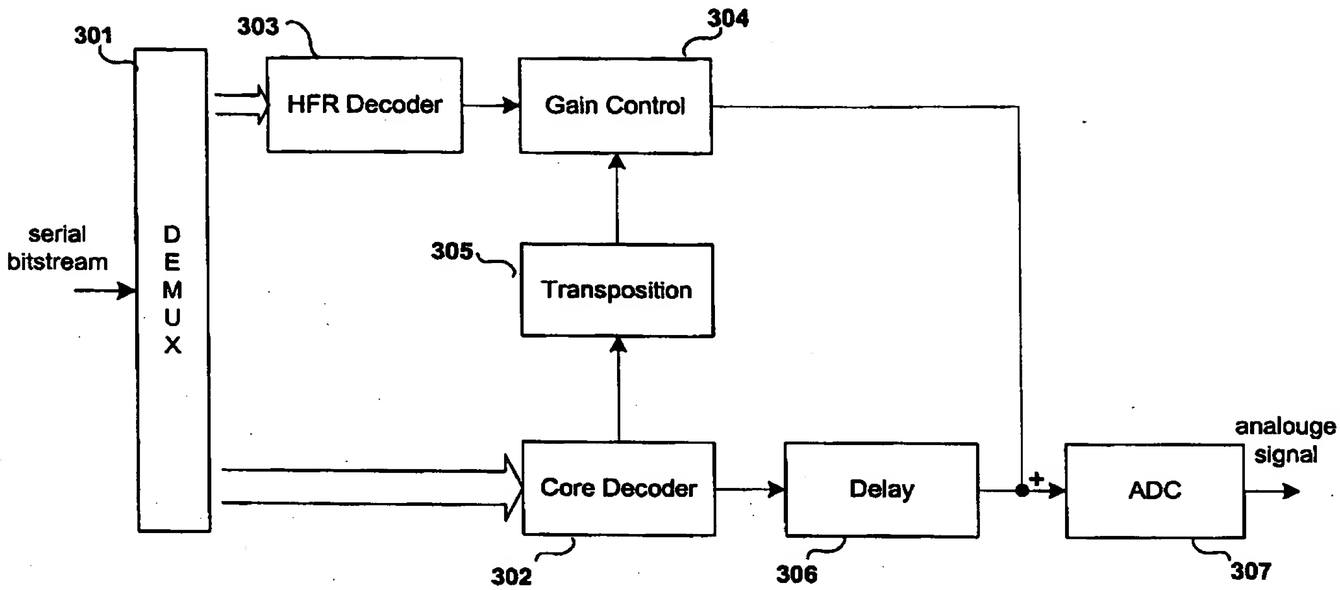

301
303
304
HFR Decoder
Gain Control
305
Transposition
serial
bitstream
D
E
M
U
X
Core Decoder
Delay
ADC
analouge
signal
+
302
306
307

Fig. 3

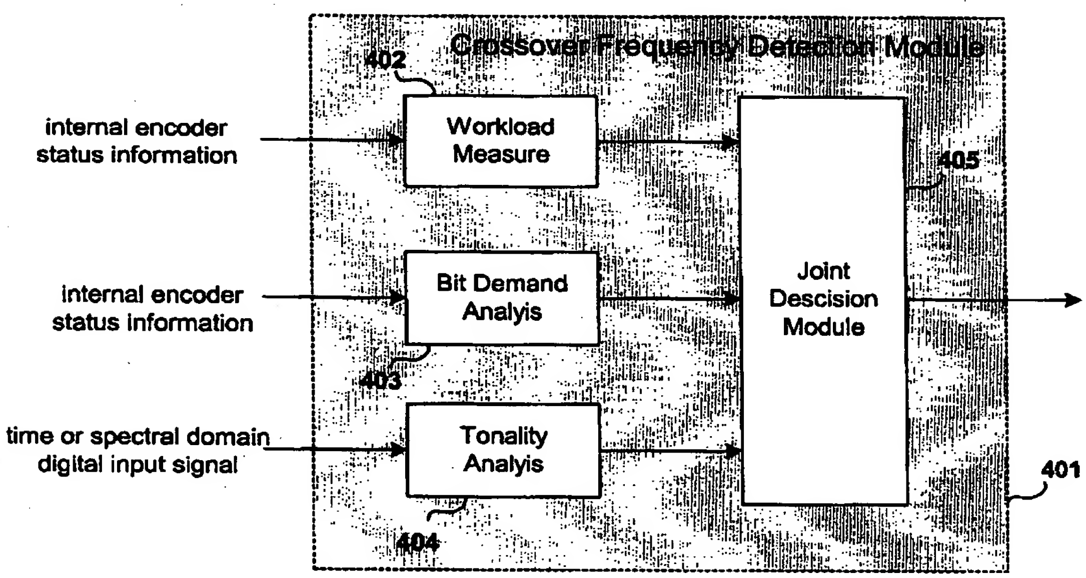
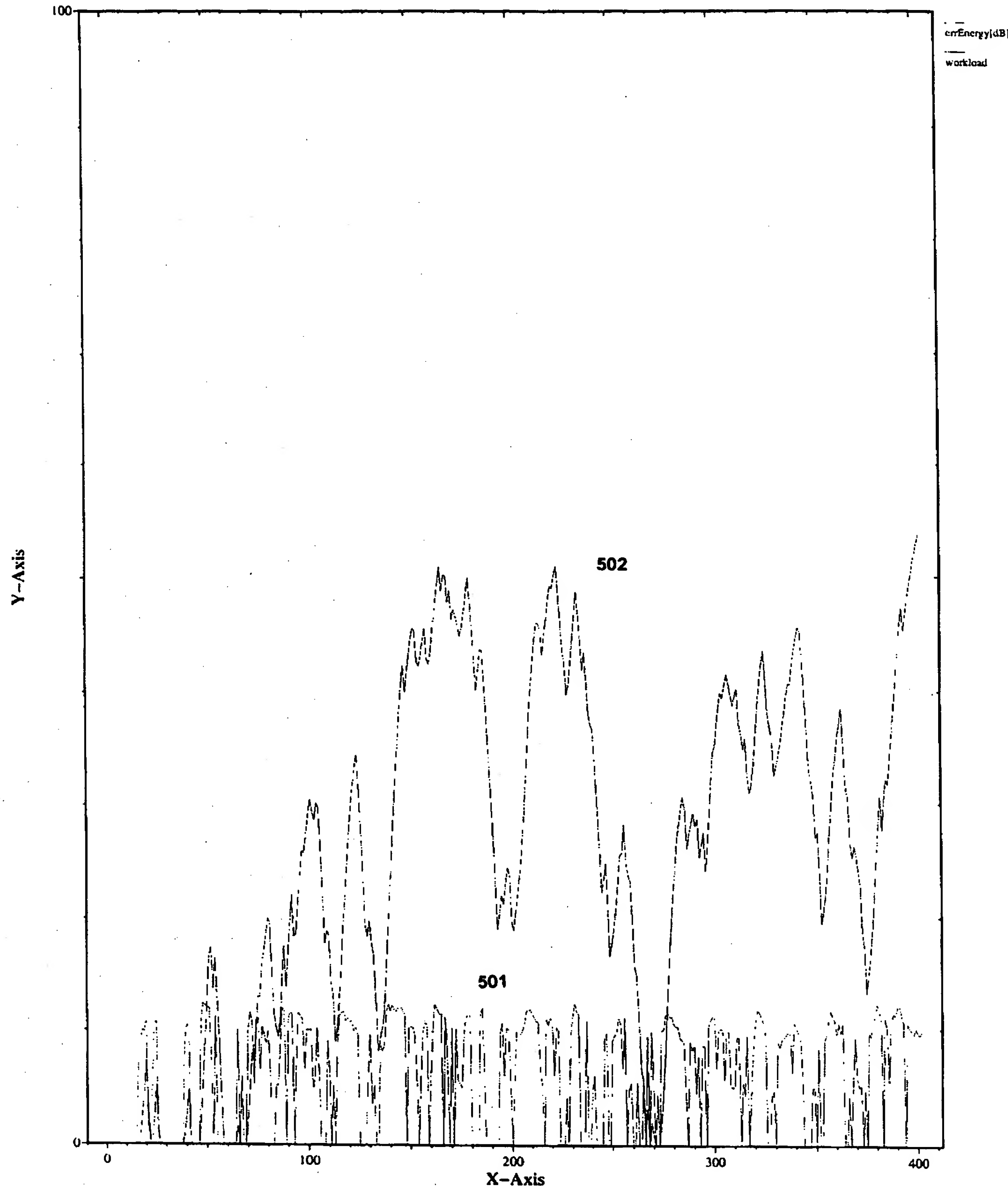
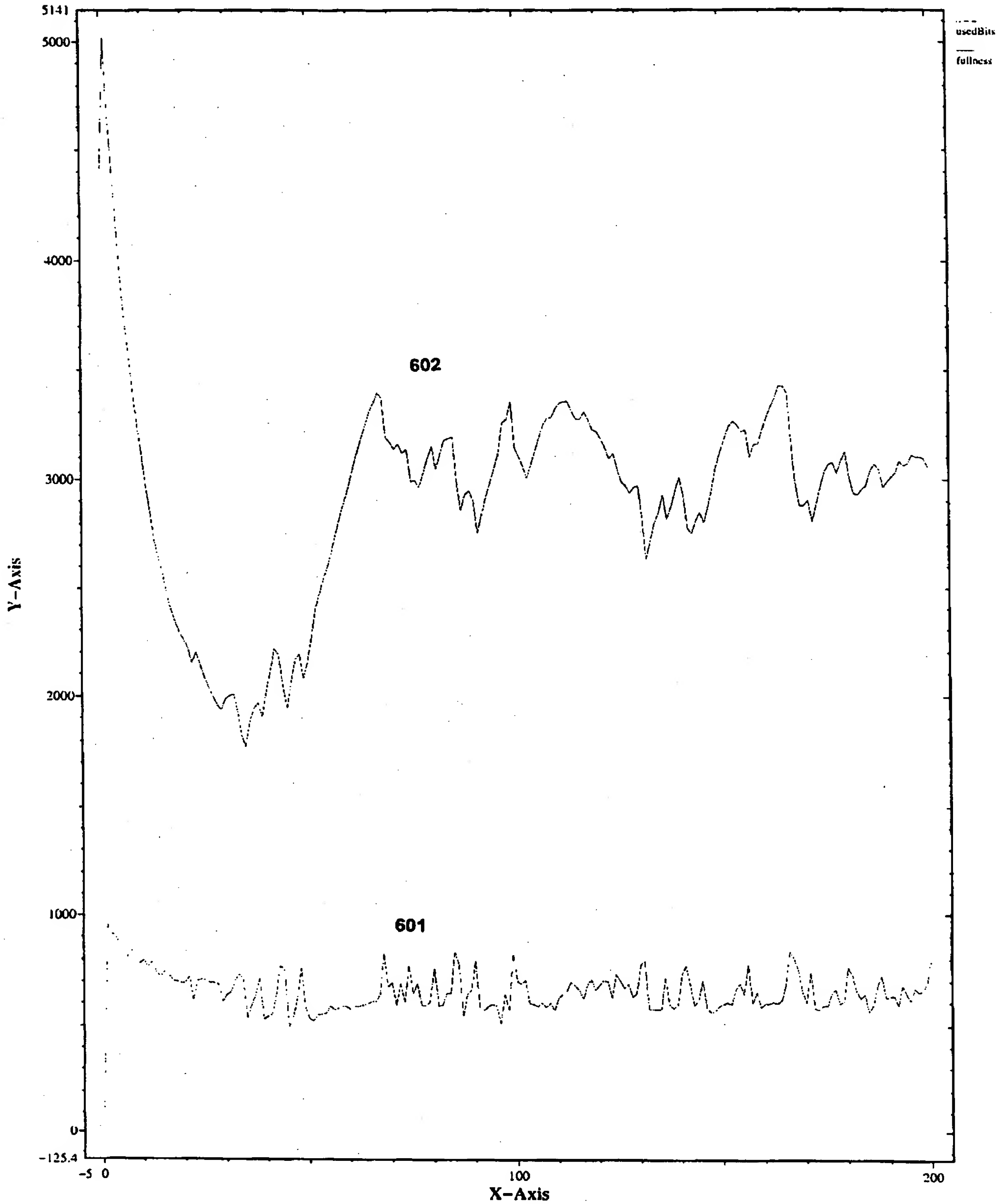
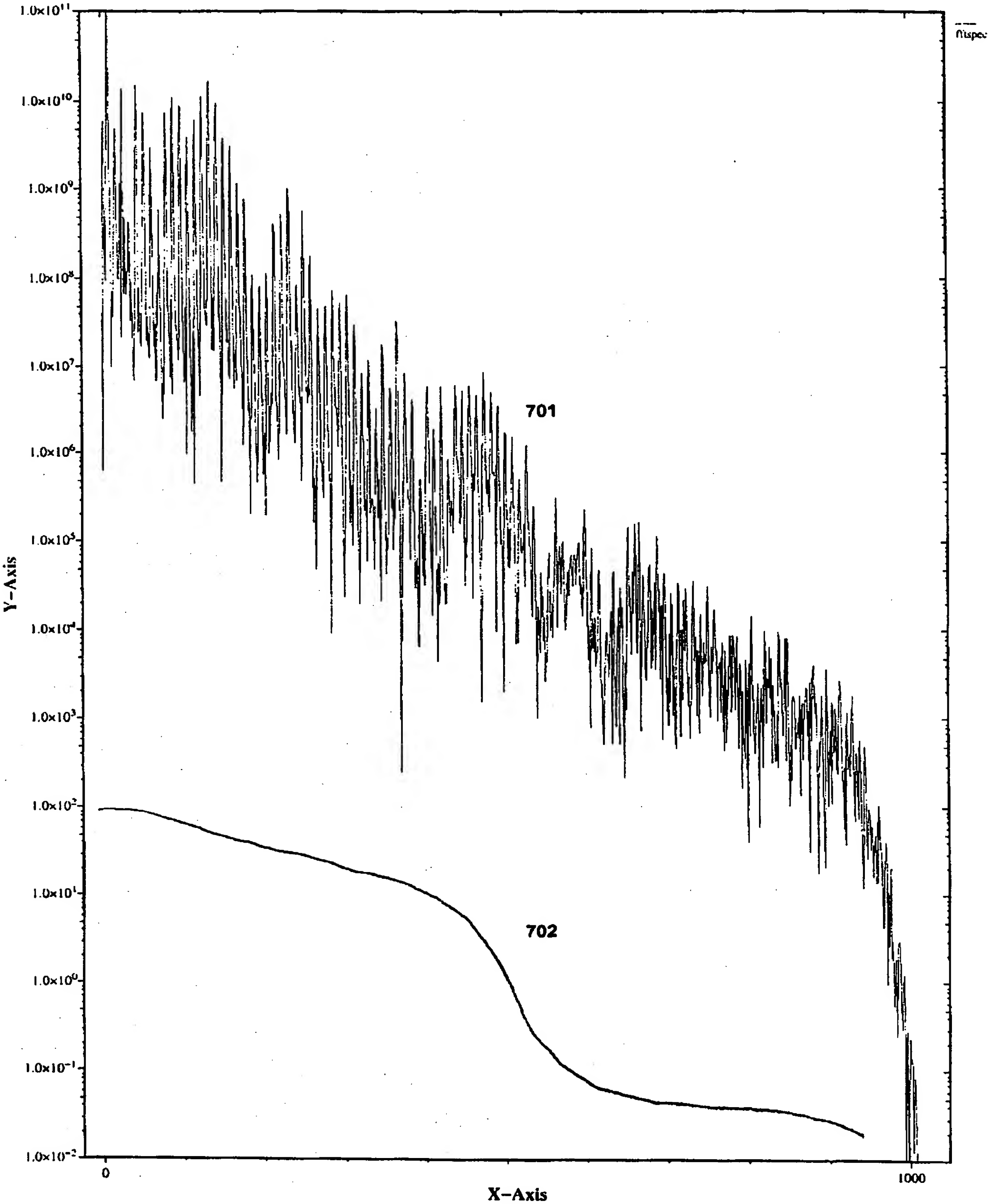

Crossover Frequency Detection Module
402
internal encoder
status information
Workload
Measure
405
internal encoder
status information
Bit Demand
Analyis
Joint
Descision
Module
403
time or spectral domain
digital input signal
Tonality
Analyis
401
404

Fig. 4


errEnergy[dB]
workload
Y-Axis
X-Axis
100
0
0
100
200
300
400
502
501


5141
5000
4000
3000
2000
1000
0
-125.4
-5
0
100
200
X-Axis
Y-Axis
usedBits
fullness
602
601


1.0x10¹¹
1.0x10¹⁰
1.0x10⁹
1.0x10⁸
1.0x10⁷
1.0x10⁶
1.0x10⁵
1.0x10⁴
1.0x10³
1.0x10²
1.0x10¹
1.0x10⁰
1.0x10⁻¹
1.0x10⁻²
Y-Axis
fftspec
701
702
0
1000
X-Axis